EPIC2019

Ethnographic Praxis in Industry Conference Proceedings

Hybrid Methodology

Combining Ethnography, Cognitive Science, and Machine Learning to Inform the Development of Context-Aware Personal Computing and Assistive Technology

MARIA CURY*, *ReD Associates*
ERYN WHITWORTH*, *Facebook Reality Labs*
SEBASTIAN BARFORT, *ReD Associates*
SÉRÉNA BOCHEREAU, *Facebook Reality Labs*
JONATHAN BROWDER, *Facebook Reality Labs*
TANYA R. JONKER, *Facebook Reality Labs*
KAHYUN SOPHIE KIM, *Facebook Reality Labs*
MIKKEL KRENCHEL, *ReD Associates*
MORGAN RAMSEY-ELLIOT, *ReD Associates*
FRIEDERIKE SCHÜÜR, *Cityblock Health*
DAVID ZAX, *ReD Associates*
JOANNA ZHANG, *ReD Associates*
* *Denotes co-first authors.*

The not-too-distant future may bring more ubiquitous personal computing technologies seamlessly integrated into people's lives, with the potential to augment reality and support human cognition. For such technology to be truly assistive to people, it must be context-aware. Human experience of context is complex, and so the early development of this technology benefits from a collaborative and interdisciplinary approach to research—what the authors call "hybrid methodology"—that combines (and challenges) the frameworks, approaches, and methods of machine learning, cognitive science, and anthropology. Hybrid methodology suggests new value ethnography can offer, but also new ways ethnographers should adapt their methodologies, deliverables, and ways of collaborating for impact in this space. This paper outlines a few of the data collection and analysis approaches emerging from hybrid methodology, and learnings about impact and team collaboration, that could be useful for applied ethnographers working on interdisciplinary projects and/or involved in the development of ubiquitous assistive technologies.

INTRODUCTION: THE POSSIBILITIES OF ASSISTIVE TECHNOLOGY, THE COMPLEXITIES OF CONTEXT, AND THE NEED FOR A HYBRID METHODOLOGY

Technology has altered everyday experience. People carry smartphones in pockets or purses and smart watches around their wrists. From light bulbs to air conditioners, today's homes are smart. Given the rate of change we have witnessed over the last decade or so, we can easily imagine a not-too-distant future that brings more ubiquitous personal computing technology seamlessly integrated in people's lives with the potential to assist people in everyday tasks. What may people want of such devices and how might we design assistive technology to give people what they want and need?

We can imagine well-timed pieces of information, a person's name, for example, discreetly delivered to avoid awkward encounters. We can imagine interventions that fit the needs of individuals in the moment, lowering the volume of background music to boost concentration. We can imagine that the playlist might be selected based on current mood or current goals. These examples highlight how interventions might be further personalized to depend on the person and the person's context. For example, one individual cooking a meal may be an aspiring chef who wants to focus on improving cooking skills and may welcome instruction and feedback, whereas another person may dislike cooking and may welcome some background music or an interesting podcast to distract from the cooking chore. Furthermore, we can imagine that it may be just as important to know when *not* to intervene, such as during a moment of deep conversation and true connection between people. In short, individuals in situations that may look alike can have very different needs. How might a device learn to understand a user and parse a user's situation, or context, to make decisions about *whether*, *when*, and *how* to provide assistance? How can such devices be designed to provide information or interventions that fit the needs of individuals in-the-moment and support how they wish to act upon their world?

At the outset of our research, we asked ourselves how we might be able to study the ways people experience complex yet everyday contexts to bring into focus the promise of future assistive technology *and* how to build it. We wanted to inform (at a very early stage) both the value such devices could offer to people in-the-moment and also how these devices might be built to parse context. In the process, we discovered the need for collaboration across disciplines and the need for a *hybrid methodology* that combines frameworks and concepts across disciplines. A single discipline's tools and approaches are likely too narrowly scoped to this new and large problem space. It requires an exploratory approach (where ethnography brings strength), combined with a focused study of internal states (where cognitive science brings strength), and it needs to be ultimately relevant to machine learning, requiring that analysis methods be informed by the types of data and data representations that machine learning will consume.

In our work, we thus drew from the disciplines of cognitive science, anthropology, and machine learning.

Cognitive science, broadly speaking, aims to characterize the nature of human perception, thought, and decision-making. Cognitive science provides methods that help us gain insight into the feats and limits of human information processing. Devices can display or otherwise share a great amount of information, and cognitive science provides the methods to understand what can and cannot be meaningfully processed by human cognition. Insights are most often gained through carefully controlled experiments in a laboratory or in a specific activity (e.g. air traffic control, Christensen et al. 2012).

Anthropology is the study of human societies, cultures, and their development and it provides us with methods that help study individuals in context. Anthropologists study and derive meaning from the everyday — observing how a range of sociocultural forces, structures, and relationships interact to form a person's experience of the world, and how that person, in turn, acts upon the world in ways that push against, reinforce, or reshape those forces. Anthropologists are experts of context, abstracting out from "thick descriptions" (Geertz 1973) of individuals to make broader reflections about human experience.

Machine learning is the study and application of algorithms and statistical models. To deliver in-moment solutions, assistive personal computing devices will need to be powered by machine learning algorithms that learn from sensor data. The challenge is that these devices need highly scalable solutions that at the same time offer strongly tailored experiences specific to individuals in context. This requires a framework that allows algorithms to abstract away from particular experiences of individuals to uncover what may be shared across individuals and situations.

Borrowing concepts or methods from all three disciplines helped us develop a more robust understanding of individuals and their contexts in ways that can support the early development of new forms of personal computing and assistive technology.

We hope hybrid methodology serves as a call for applied ethnographers to adapt their methods, deliverables, and ways of collaborating for greater impact in this space. Traditionally, qualitative data and user research are used early in the product development life-cycle to identify and scope use cases for the product and then late in the product development life cycle to gather user feedback on prototypes, finished products, and product features. Qualitative data rarely informs machine learning problem formulations or cognitive science experiments. We go beyond this traditional model toward deeper collaboration. Ethnography is no stranger to hybrid approaches — for example, anderson et al. (2009) have explored the combination of qualitative research with data mining into "ethno-mining [...] a hybrid, not a 'mixed method'; it is two elements that cannot be separated out [...] [yet] traces of each of the ingredients can still be seen - the same ethos of ethnography (open-ended, co-constructed, holistic field research) integrated with the empirical and analytical capacities of quantitative data mining" (anderson et al. 2009, 125). Applied social scientists have also been exploring the blending of "big data" with "thick data" (Bornakke and Due 2018) and outlining approaches like "Contextual Analytics: a project process for uniting data analysts and social scientists under the mandate of building more effective and credible algorithms" (Arora et al. 2018, 225). Our work hopes to carry this thinking forward.

In this paper, we outline our approach to the early development of assistive technology. In the process, we share how our hybrid methodology allows us to answer novel research questions and how it supports the development of new products. We doubt that we would have been able to achieve these results had we not all stepped beyond the comfort of our disciplines, finding ourselves in a collaborative balancing act: considering tradeoffs between the practices of one discipline and another, between the structure of the lab and the openness of the field, and between the different definitions of what data is, how it can be analyzed and processed. This paper outlines a few of the data collection and analysis approaches emerging from hybrid methodology, and learnings about impact and team collaboration that are useful for applied ethnographers working on interdisciplinary projects, particularly those involved in the development of ubiquitous assistive technologies.

SITUATING HYBRID METHODOLOGY: ABOUT OUR STUDY

An interdisciplinary team combining Facebook Reality Labs and ReD Associates researchers with expertise in anthropology, cognitive science, and machine learning sought to understand the human experience of performing tasks in everyday contexts, to inform the early development of context-aware assistive personal computing technology. We studied *experience*, or the subjective moment-to-moment internal states of our participants, with a particular focus on the experience of mental effort. And we studied *context* itself.

It is our belief that by taking the broadest possible view of context, we can build ubiquitous devices that are truly useful partners to people, enhancing their agency through a smart and sensitive parsing of the fullness of their experience. Linguistic anthropologists Alessandro Duranti and Charles Goodwin define context as "the frame that surrounds an event and provides resources for its appropriate interpretation" (Duranti and Goodwin 1992, 3). Context in its broadest sense includes not just spatial context (a physical environment), but also layers of social context, personal/psychological context, and temporal context. In order to study contexts in-situ, we began by de-constructing a context into its component parts for observation. Our categories were similar to those used in Activity Theory, a framework from the social sciences which acknowledges how the physical environment, social dynamics, cultural norms, objects, and the individual mind are interconnected in an activity (Engeström et al. 1999; Kaptelinin and Nardi 2006; Nardi 1996; Roth 2004).

We wanted to ensure we were attuned to the different elements of a context while in the field (and how those elements interacted), and careful not to collapse context to the sum of its parts in-the-moment, so our field guide included prompts to systematically observe each component we defined. We used a range of theories to shape our understanding of the different components of context, mixing theories in perhaps 'low fidelity' or bricolage ways that focused on drawing out and combining the aspects of each theory most helpful to our research question (Cury and Bird 2016). For example, we brought theories about the built environment's impact on human experience (Goldhagen 2017) in conversation with findings on the physical environment's impact on mental effort (Choi et al. 2014) and with theories about how the passage of time can be perceived through distinct spaces and tasks (Ingold 1993). We considered how people's movements and ways of seeing are socially constructed and learned (Mauss 2009; Grasseni 2007), together with findings on the effects of visual training on how people make observations (Braverman 2011). All of these together helped us to build a multidisciplinary understanding of context to explore in the field.

We met with eighteen participants from Seattle, New Jersey and New York — eight females and ten males, ages ranging from 25 to 54, with diversity in ethnicity, occupations, home types (e.g. apartment, house), and living arrangements (e.g. single living alone, roommates, couple without kids, couple or single parent with kids). The researchers disclosed the identity of the organizations conducting the research and the high-level aims of the research to each participant prior to the participant's voluntary consent to join the study. Participants were compensated for their involvement, were informed that they could withdraw from the study at any point, and were given opportunities to ask questions about the study and its methods.

Two researchers met with participants for a full day session in their homes and communities, accompanying them during their daily routines. Drawing from the sensory ethnography guidelines of anthropologist Sarah Pink (2009), researchers conducted participant observation in which they were attuned to their own sensorial experience of the spaces they were in with participants, and in which they asked participants to reflect on both the abstract and sensorial aspects of the activities they were doing and the objects they were using. Researchers conducted semi-structured interviews about participants' home life and personal history, and various exercises to map what tends to occupy their "headspace" on a given day, their relationship to technology, and their social ecology. The research centered on systematic observation of two focal activities, followed by in-depth discussion with the participant after each activity. For the focal activities, all participants cooked a meal in their kitchen and performed a second goal-oriented activity of their choosing (e.g. doing laundry), mostly in their homes. We strove for variability in how participants generally felt about the two activities based on self-reporting ahead of time along key factors such as whether they found the activity enjoyable.

Prior to the activities, participants completed a training session to understand the concept of mental effort (from the field guide: "*The amount of mental activity that is required while you're doing some task or tasks. This mental activity can involve thinking, deciding, calculating, remembering, searching, etc.*"). They were provided with a definition and analogies, a numerical scale to use during the activities, and a series of exercises to practice reflecting on mental effort and ensure comprehension of the concept. During both activities, participants were recorded using a wide-angle camera to capture the physical context of the activity. In the cooking activity, participants also wore a head-mounted camera that captured the activity and context from a first-person perspective. During the two activities, one researcher recorded ethnographic field notes, while the other probed the participant to report their mental effort periodically and systematically. The visual recordings as well as the self-reports were used during the de-briefing discussions with the participants.

We captured a variety of qualitative and quantitative data, including thick ethnographic field notes, descriptive mental models and maps (e.g. of "headspace," technology use, social ecology) drawn together with the participants, repeated numeric mental effort scores with verbal descriptions of contributing factors, and high-resolution video data of a participant's context and first-person perspective. We analyzed these data in a similarly varied way upon return from the field, drawing on approaches from the researchers' "home disciplines." It is from these data and analyses that we generated insights, abstractions, and data labelling protocols for parsing context, that have now advanced into the work of machine learning and cognitive science teams (see Jonker et al. in review, for selected findings).

This project — with its ambition to understand human experience of context for technology development — required a constant dialogue across disciplines that study dimensions of experience and context. The project required combining methods, frameworks, concepts and ultimately data from anthropology, cognitive science, and machine learning (alongside philosophy, linguistics, and journalism). It also required applied ethnographers to push the boundaries of what constitutes data, an insight, and an output of research, to be relevant. What follows is an outline of a hybrid methodology that may guide interdisciplinary teams to better collaborate, and for ethnographers to find new applications of their work.

HYBRID METHODOLOGY RESEARCH: DEVELOPING RESEARCH METHODS

Interdisciplinary projects have an interdependence of methodologies, and each method gets a little bit "sullied" as it moves out of its intended disciplinary realm and into a hybrid space. For instance, when ethnography moves to the semi-structured environment of the participant's kitchen that is now set up with conspicuous cameras and two researchers (one of whom is asking scale-of-one-to-nine questions systematically every three to five minutes), "pure" participant-observation is, arguably, not happening. If the represented disciplines' experts each feel slightly uncomfortable with the imperfection (or slight irreverence) with which their methodologies are being deployed, the team may actually be in a good place. The emphasis is on triangulation and testing, with the ultimate deciding factor for choosing and melding together methods being: what is most in service of answering the research question?

What follows are two research methods, drawn from our study, that combine approaches from different disciplines to help answer the research question "how do humans experience everyday activities in daily contexts?" to inform the development of new personal computing and assistive technology. For researchers with similar research questions, the two methods described here may be directly relevant. For researchers with different research questions but a similar interdisciplinary team set-up, the methods described here may serve as an example for how other methods, from other disciplines, may be hybridized to suit the needs of the research question.

The first method we describe, experience sampling in participant-observation, combines an approach from social psychology with ethnographic research, to gather data on the experience of context in-real-time. The second method we describe, reconstructed narratives with video playback, involves the active role of the research participant in reflecting on their experiences using video footage, to gather data on internal states that would otherwise not be gleaned from researcher observation alone.

Experience Sampling in Participant Observation

How do researchers capture a person's momentary experience in a way that lends itself to systematic, multi-disciplinary analysis? First, measurements should be captured in-the-moment, to give us access to the often-transient experience during a task, and to avoid biases in retrospective recall (Redelmeier and Kahneman 1996). Second, the protocol itself has to be relatively non-intrusive, to not affect the person's experience in the moment. And third, the measurements should be simple, to avoid selection bias and ensure meaningful responses from all participants.

One approach for doing this is experience sampling (or event sampling), a widely used method in social psychology (e.g. Reis and Gable 2000; Larson and Csikszentmihalyi 2014) and cognitive science (e.g. Kane et al. 2014; Killingsworth and Gilbert 2010) to consistently elicit subjective thoughts, feelings, and behaviors in the moment. For example, a researcher studying adherence to a new health habit might have research participants install an app on their phones that pings a prompt to them twice a day, asking for a reflection about how tired they are feeling or about whether they completed a routine. Experience sampling allows researchers to capture a representation of experience as it occurs, and to analyze patterns and relationships as they unfold over time. The repeated measurements are collected in different contexts and during various tasks and sub-tasks, enabling researchers to unpack and disentangle the complex contextual factors affecting subjective experience. Because experiences are captured in the moment, rather than after-the-fact, participants are less likely to suffer from memory bias. In retrospect, people tend to overestimate the difficulty of certain tasks and the amount of energy applied to solving these (Schmeck et al. 2015). Furthermore, experience sampling is a validated tool that enables researchers to compare results across study sites (such as a lab versus a naturalistic environment).

Compared to laboratory experiments, experience sampling methods have the advantage of collecting data in the participant's everyday contexts. This allows researchers to observe thoughts and feelings as they occur during everyday activities that can be difficult to recreate in more controlled settings. Indeed, in a study of mind wandering, researchers discovered significantly higher frequency of mind wandering in daily life than is typically seen among participants in laboratory experiments (Killingsworth and Gilbert 2010). Further, it allows researchers to understand not only how participants experience certain tasks, but also how much mental energy they invest in the task — an aspect that is crucial to development of assistive technology, as described in the introduction.

However, experience sampling methods place heavy demands on researchers and participants alike, and as we found, require careful instruction to ensure that all participants are comfortable reporting their answers. When conducting experience sampling in the context of ethnographic observation, it is important to first build rapport with the participant. For example, we first met with Marcus, one of our participants, over lunch before he attended his afternoon lecture, we met with him again afterwards and in total spent several hours talking more broadly about his daily life, interests, history, and social ecology, and observing his surroundings (his favorite food stall, his commute home) with him before any experience sampling took place. When it came time for Marcus to cook (he does batch cooking once a week to unwind from the stresses of medical school), we first took a pause from his routine to train him on experience sampling. We took a candid tone throughout ("this might seem a little goofy but...", "we're going to be annoying flies on the wall buzzing every so often with a question...") to mitigate the "experimental" feel of the method, which is at odds with the everyday "deep hanging out" (Geertz 1998) feel of ethnography. It is important that the research participant ultimately feels familiar and comfortable with the method (and with being interrupted every so often with a question).

Experience sampling designs come in many shapes and forms. Time sampling probes the person at fixed intervals. Random sampling probes the person at random intervals throughout the activity. Event sampling probes the person during particular events. The rule-based approach of time sampling guarantees systematic data capture, but lacks the flexibility to capture the influence of interesting events that often lack clear beginnings and ends. When conducting experience sampling in an ethnographic context, a mixed approach can account for the open-ended nature of everyday contexts with its interruptions and surprises. We decided on a mix of time- and event sampling, in which we systematically probed the participant every two to four minutes during brief moments of downtime (e.g. pausing after draining the noodles), but encouraged the researcher to conduct additional probes whenever interesting events, as determined by the researcher, occurred (e.g. a paper towel accidentally getting caught on fire).

Experience sampling design includes not just how often sampling occurs, but also what is asked of participants. Because of our research question and project goals, we used a subjective mental effort rating scale: participants are asked to report answers to the question "How much mental effort did you invest?" on a 9-point Likert scale ranging from *very, very low* to *very, very high* (Paas 1992). While having repeated quantitative measures proved to be very valuable for our project, the Paas scores (what we will refer to as "mental effort scores") themselves gave us limited insight into the contextual factors that shape a person's experience in a given moment. We needed more clues to understand how numerous factors influence a person's mental state, including task complexity, engagement, emotions, social environment, and so on. Therefore, we asked participants to explain, in a few simple sentences, what they were thinking of, or other things that preoccupied their minds, after having reported their mental effort score — informally calling the qualitative adaption "Paas + why."

There are many ways a participant can answer "why," and it is important to strike a balance between providing room for freeform reflection and providing structure for reflections that can be compared across participants. Matthews et al. (2013) and Helton and Näswall (2015) uncovered three primary dimensions of so-called stress states (transient states during a task that permeate conscious awareness): engagement, distress, and worry, mirroring the "trilogy of psychology," motivation, affect, and cognition. We developed the training material described above, to familiarize participants with experience sampling, such that it trained the participant not only on how to use numeric scales, but to begin to develop a sensitivity for breaking down their experience into component parts — asking them to reflect also on task difficulty, engagement, and feelings toward the task in their open-ended answers to toy problems we gave them as part of the training. We encouraged them to later consider these aspects when giving their "why" answers to the mental effort scores once cooking commenced. We used a modified version of the Weekday problems (Sweller 1993; Van Gog et al. 2012) — for example, "*Suppose tomorrow is Monday. What day of the week is five days after the day after tomorrow?*" (Schmeck et al., 2015) — that we altered to vary not only in difficulty (high, medium, low) but also in engagement (artificial high incentive, artificial low incentive). We imposed this variability in both difficulty and engagement to allow the participant to reflect on the *choice* as to how much effort to invest in a task. This is meant to mirror the fact that in a real-life context the difficulty of and the participant's engagement in the activity will vary in ways we cannot control but in ways we want to *understand.*

The protocol was tested and refined during the initial research phase. One key learning was that some participants struggled to disentangle emotion from cognition (e.g. watching a movie may be very emotionally moving but require very little mental effort to comprehend or watch, unless it is in a foreign language or it causes someone to mind-wander and reminds them about a to-do list). This led to additions to the training protocol to help people disentangle the two dimensions while signaling that both dimensions are equally important. For instance, we asked participants to establish a "benchmark" by providing a previous experience in their own lives that they would consider a mental effort score of 1 and a mental effort score of 9, after they were trained on the concepts and toy problems. This allowed researchers to both correct any misunderstandings of the concepts and also to contextualize the participants' later scores with other aspects of their lives, for richer qualitative data. We encouraged participants to report on emotions when asked "why" for their scores. But convincing participants of the researchers' equal interest in emotion was complicated by the fact that we had no quantitative approach for measuring emotion as we did for effort. Some participants interpreted this difference to imply that their emotions were of secondary importance. Future work might benefit from developing such a scale and deploying it side-by-side with the mental effort scale in everyday contexts (see Fraser et al. 2012 for connections between emotion and cognitive load, and Lottridge et al. 2011 for conceptualizations and measurement strategies for emotion).

Combining experience sampling and mental effort scoring with ethnographic participant observation requires compromises to each of the methods. In this project, it required rapport-building (in part through the researcher's candid self-reflection on the strangeness of the method) prior to experience sampling, and a mix of time- and event sampling using the researcher's discretion and including room for open-ended reflections of "why" in addition to scores. Understanding mental effort required training to tease apart different aspects of everyday experience like types of stress, emotion versus cognition, and the choice to engage in a challenging activity at all. These are all aspects that might be controlled for in a lab, but which we tried to capture and record the variability of in everyday contexts. Experience sampling in participant observation also created a setting that was more structured and with more interventions on the part of the researcher than in a classic ethnography. In these ways, disciplinary experts found themselves uncomfortable, and found the data less pristine than they would have hoped, but ultimately the fieldwork collected qualitative and quantitative data that explored context and human experience from various angles and with aspects that each discipline alone would not have been able to capture.

Reconstructed Narratives with Video Playback

How does a researcher break down, in moment-by-moment sequence, another person's experience? The researcher can observe someone in real time, but that does not explore interiority (e.g. what is our participant Marcus deciding between as he's stirring the pot of noodles? What caused him to pause for so long by the window?). Researchers could interrupt that person at a steady cadence to probe deeply at interiority beyond the "Paas+why" experience sampling described above, but that would introduce an "observer effect" distortion. The in-depth questioning could prevent the participant from entering important and common subjective states such as "flow" states (Csikszentmihalyi 2008) or mind-wandering (Smallwood and Schooler 2015) that would otherwise typically occur when the participant is in the everyday context and that would be helpful for the researcher's understanding of what assistance, if any at all, a person might need in that context.

As our team puzzled over this problem, we began to look to what was in retrospect one of the techniques of narrative journalism: the reconstructed narrative interview (Menkedick 2018). Journalists who specialize in telling narrative stories deeply rooted in one "character's" experience rapidly learn the value of revisiting with an interview subject a particular event again and again; each visit adds a new layer of depth, and helps the journalist to recapture what it was like to live through that event. In designing our research, we settled on a version of this technique as a method to probe participants' experience of cooking in a way that was both deep, yet unobtrusive: we would allow the participant to perform his or her task with no questioning beyond the mental effort scores asked every two to four minutes, and only *after* the cooking was complete would we engage the participant in an interview of approximately 60 minutes (sometimes longer) to immediately reconstruct, with as much fidelity as possible, what the interior experience of the just-completed task had been, particularly during a few moments of interest informed by steep changes in the mental effort scores they reported. For each participant we did this process twice, after each of the two activities. Crucially, we scrolled through the just-captured first-person video of the participant doing the activity during the interview to guide the questioning.

With in-situ fieldwork, researchers have an advantage over the journalist, as well as a disadvantage. The advantage is presence. Journalists are rarely physically present during the "scenes" or moments they later seek to reconstruct in their subjects' lives. By contrast, researchers in-situ are able to quietly observe and take notes about the scenes they will shortly try to reconstruct. Research can be set up to have the further advantage of being able to conduct the debrief interview *immediately* following the task; a narrative journalist often is piecing together events that date back years or even decades. The disadvantage is that researchers are seeking to reconstruct the experience of essentially banal events (e.g. doing laundry), and on a more minute time scale than a journalist would try to explore (e.g. returning the shirt to the ironing board just when it seemed like the shirt was done getting ironed). Very seldom does a journalist attempt to reconstruct how a person's experience shifted across the course of a second, and never would a journalist expect a subject to remember with any fidelity the precise order in which the subject executed essentially banal tasks, like whether salt was added to a broth before pepper, and why.

Video footage can be used to overcome this challenge. In our study, we decided to play back to participants the video that had been recorded of them performing the cooking task using a head-mounted camera. (During the second activity of the participants' choosing, there was only an in-room camera recording the activity. We decided on this approach in case the head-mounted camera proved to be too disruptive for the participants' experience, but participants reflected that for the most part they forgot about the head-mounted camera after a few minutes of cooking.) This video, if instantly replay-able, serves as a kind of memory prosthetic to assist reconstructive narrative interviewing; the first-person perspective of the camera view further helps the participant relive the experience of the hour before. For instance, vision darting from one ingredient to another could help the participant viscerally remember a moment's indecision over how to proceed with a recipe. (We also realized that participants were much more comfortable watching first-person video of themselves than room-camera video of themselves that often made participants feel self-conscious.)

The reconstructed narrative with video playback can take longer than doing the activity itself, but it is this time investment that allows for deep probing into what would otherwise remain unseen or untranslatable to the researcher — a furrowed brow, a pause, a chuckle. Moments that are apt for deep discussion can be selected by both the researchers, looking back on their notes, and the participants, recalling something they had thought about but didn't say aloud at the time. Following the cooking task, we sat down with the participant and spent about an hour reviewing moments of special interest with the participant. Moments of interest were chosen at the researcher's discretion, but often involved spikes or significant fluctuations of mental effort as recorded from the mental effort score self-reports, moments of clear task-switching, moments of interruption, or moments the researchers had trouble deciphering. The researchers also allowed the participants to highlight moments that to the researchers seemed uneventful but where internally within the participant there was a lot of activity. For instance, one participant Haley noted that when she was waiting for the tofu to brown she was reminded of a reply she was waiting on from a love interest. The researchers soon discovered that to thoroughly explain everything that influenced the participant's experience during a moment of high complexity — even if that moment only lasted 30 seconds — could easily take 20 minutes of exhaustive probing through repeated playback of the video clip.

To give one example: one researcher witnessed a participant, Daryl: 1) have a dialogue with his wife about a task related to their young daughter's pajamas, 2) make a note about this task on a nearby whiteboard, 3) rapidly decide to execute the task immediately instead, thereby abandoning his borscht recipe for the moment, 4) quickly visit different drawers in his daughter's bedroom (captured for the ethnographer only due to Daryl's wearing a head-mounted camera, as he had darted away from the kitchen at this point), 5) visit a drying machine to grab a pair of pajamas, then 6) finally return to his borscht. Puzzling out all of these decisions, and the sub-decisions within these decisions, was a laborious (if joyful) task for the researcher, necessitating digressive interviews about the state of Daryl's relationship with his wife, his young daughter's aversion to wearing pajamas, and a history of the participant's forgetting to execute tasks placed on the family chore-board. The entire video clip lasted perhaps just 30 seconds, but the exhaustive and fully explanatory account of the *meaning* of it ran for several hundreds of words.

This method of narrative reconstruction using first-person point of view video playback builds on participatory ethnographic video practices (see for example Pink 2007, 103-115), and places emphasis on the research participant's role in interpreting and making sense of their own experiences, rather than leaving the interpretation and sensemaking to the researcher alone upon return from the field (as may often be the case for the ethnographer) or from the lab (as may often be the case for the cognitive scientist). As anthropologist João Biehl writes, "How can the lives of our informants and collaborators, and the counter-knowledges that they fashion, become alternative figures of thought that might animate comparative work [...]? [...] As anthropologists, [...] we are challenged to listen to people — their self-understandings, their storytelling, their own concept work — with deliberate openness to life in all its refractions" (Biehl 2013, pp 574-6). This is perhaps another way in which hybrid methodology seeks to push the boundaries of research — by bringing participants more actively into the sensemaking process — and future work might benefit from developing this aspect further. Providing research participants more opportunities to articulate their internal states, including what they need and what they *don't* need, rather than assuming or inferring from observations alone, seems particularly important for determining the relevance, helpfulness, and boundaries of an assistive technology in everyday contexts.

HYBRID METHODOLOGY ANALYSIS: ANALYZING DATA WITH COMPLEMENTARY APPROACHES

Because of the mix of methods combined in research, hybrid methodology generates a substantial amount of data of different types (e.g. numerical scores, observational field notes, images, video recordings). Given the wealth of data collected, many analysis strategies are possible in order to make sense of that data. The interdisciplinary team needs to choose which means of analyses to prioritize and combine in ways that best serve the research question (rather than in ways that best serve each discipline). In the case of complex research questions (e.g. what is the human experience of context?), conducting complementary analyses that make simultaneous entry points into the data allows the team to explore the research question from different angles and to revisit the data later on as distinct disciplines follow particular tracks to explore a sub-component of the research question more in-depth.

In this section we present a selection of complementary analyses that we conducted, which combined qualitative and quantitative approaches. These analyses are part of a larger pattern recognition or "Sensemaking" process (Madsbjerg and Rasmussen 2014; also described in Hou and Holme 2015), in which teams use "bottom-up" data-driven approaches (i.e. based on what we see in the field) alongside "top-down" theme-driven approaches (i.e. based on the themes we sought to explore at the outset and questions we needed answers to). In our case, we wanted the results of the analyses to help inform the early design of new assistance experiences, the research agenda for further studies (in a lab and in the field) based on new questions emerging from the work, and the early development of infrastructure for new assistive personal computing technology.

Structured Storytelling and Qualitative Data Clustering

How do teams ensure that all researchers are familiar with the details of the raw data and have a shared starting point, particularly when each researcher met with only a subset of participants? How do we enable researchers to discern themes across distinct moments in the field? We took what we informally called "structured storytelling" as our starting point in analysis: a discussion centered on each of the research participants, led by the researchers who met with that participant, and structured around key questions and instances from the field that the team wants to systematically and consistently probe for details. This ensures that human voices and experiences are top of mind — the participants are not abstracted as "Subject A" or as data points on a graph, but instead as individuals with names (we used pseudonyms to protect identity). It also ensures all team members have a shared grasp of the details and particularities of the fieldwork, from which (when those details are compared, connected, and abstracted) insights tend to emerge.

In the discussions, the team focused on concrete moments observed in the field — Dina doing laundry, or Mitchell tending to his indoor garden. This involved re-watching video footage around moments that were quantitatively interesting because the participant reported a very high or very low mental effort score, and moments that were qualitatively interesting because of an ethnographically rich observation (e.g. a moment the participant identified as meaningful upon reflection after the activity was done or a moment the researcher noticed as having many contextual dimensions at play). The purpose of structured storytelling is to interrogate the raw data with pertinent lines of questioning that help the team to interpret what happened in the field. Some of the questions we asked as a team included, "What dimensions of the context were especially relevant for the individual in this moment?" "What type of information was the individual engaged with?" "What other moments from the field, from this participant or other participants, might be similar to this one, and why?"

Structured storytelling stems from grounded theory, a methodology used in sociology and anthropology to generate theories based on systematic analysis of qualitative data rather than using data to confirm or refute a hypothesis, or building research around an existing theory (Glaser and Strauss 2017). Structured storytelling, as described above, generates interpretive descriptions or reflections that the team members then write down individually (e.g. on post-it notes or note cards) and aggregate collectively. This content, in turn, leads the team to do qualitative data clustering, which entails making further sense of the interpretive descriptions by grouping them into thematic buckets based on commonalities. These buckets are then analyzed, connected, and compared to develop working theories or insights. The development of these theories requires a constant "zooming out and in" — once there is a potential insight (i.e. a working theory that explains observations from the field), it is necessary to go back to the raw data itself to collect other moments (e.g. moments that corresponded with similar mental effort score, or moments that were ethnographically rich) that support, nuance, or challenge the proto-insight, for its refinement.

A team can tell whether or not the structured storytelling and qualitative data clustering are going in the right direction if there is a certain productiveness to the hypotheses or proto-insights — these are helping to reframe or give new meaning to moments in the field not otherwise considered, are leading to other proto-insights, or are providing structure and groupings in an otherwise fragmented data set of moments from the field. The purpose is to develop high-level insights that address the project's research questions and ambitions — in our case, about the role of different dimensions of context on a person's experience that then informed the abstractions we developed for a data labelling protocol, described in the Impact section. The abstractions we developed (which we refer to in this paper as Abstraction Set A and Set B and which can be thought of as an early framework that *informs* the later framework the assistive technology itself might eventually use) were based on the strongest patterns in our qualitative data clustering exercises and the relationships those patterns had to the quantitative analysis we will now describe.

Quantitative Analysis of Ethnographic Data

To allow machine learning models and cognitive science research to benefit from insights derived from qualitative analysis, we need to also find complementary quantitative methods for data analysis. How do teams work quantitatively with data captured in ethnographic research? Quantitative analysis of ethnographic data entails developing an approach to data processing and graphical representation to best serve the team's goals. We had three learnings that could be useful for teams doing this type of work: First, if in doubt about what type of quantitative analysis will prove useful, the team should develop multiple initial representations of the same data to enable a variety of early insights. Second, the team should seek ways to compare data points consistently and systematically even when individual research participants' experiences or real-world contexts and interpretations of tested concept are highly variable. Third, the team should explore connections between the quantitative and qualitative data to better understand the results of the quantitative analyses and address project goals (e.g. in our case going back to the thick descriptions associated with extreme mental effort scores to find other patterns in this data).

One of our goals was to obtain generalizable patterns about mental effort from the mental effort scores. The challenge is that, given the uncontrolled situations we were studying, the mental effort scores were generally not comparable across participants because of variable real-world contexts and because of individual differences in how participants interpreted the mental effort scale. This is a common problem with all self-report scales. For example, one participant never gave a maximum score of 9 (always hovering around 6s or 7s at the extreme), but her qualitative *description* of a moment was very similar to another participant's description for a 9 score. This left us with an interesting question: Can mental effort scores be compared across different activities for the same participant, and across participants?

We plotted the mental effort scores for each participant's two activities first in box-and-whisker plots, which allowed us to visualize the median mental effort score the participant gave for that activity, as well as the upper and lower quartiles of that median and the upper and lower extremes (moments when the participant gave a really high score or a really low score, outside of the norm of scores they were otherwise providing). We were able to contextualize these plots with what we knew qualitatively about each participant, to identify patterns in how each participant "typically" scores mental effort (e.g. Marcus loves cooking and it's easy for him, whereas he doesn't enjoy studying and finds the material difficult, but there are relative "extremes" in each activity, with distinct needs, and those might have similarities to another participant's, when we begin to abstract out through the qualitative data clustering).

In order to paint a picture of how each participant's mental effort reports shifted over time, we made another set of mental effort score plots with score values on the y-axis and time on the x-axis. This provided a "story arc" of how an activity unfolded in terms of mental effort from start to finish, which we could then contextualize with qualitative data (e.g. Dina did laundry late in the day feeling rushed to get it done while the food was cooking, so perhaps that's why the "arc" of the activity looks the way it does). We could also compare the mental effort score arc with what we knew from the reconstructed narratives in the field (e.g. when Haley's scores were low during a banal moment in cooking, we knew she was thinking about her romantic interest and about her work responsibilities). We were able to assess where our ethnographic observations differed from or aligned with the mental effort scores, and understand how two participants' needs, when compared, were distinct even when they each gave a score of 9 during a moment when they were each cooking.

To better visualize the set of high mental effort "outliers" (the particularly rich moments from a cognitive science point of view) and identify clusters (similar patterns) between participants, we calculated the mean and standard deviation of the mental effort scores across both activities for each participant, plotted in temporal sequence (how the mental effort scores changed over time for each participant). High outliers were defined as those that fell in the top 10% of the distribution for a participant. Because we had qualitative notes accompanying each score, we were able to interpret and theorize about why a moment was an extreme high or low score, for that individual, and find patterns among the "why's" behind the relative extreme scores. This data informed subsequent analyses conducted by the cognitive scientists on our team (Jonker et al. in review).

Multiple forms of analyses are possible on, and can enrich our understanding of, a hybridized data set, to provide more directional outcomes. Together these approaches set our team up to explore further cognitive science questions around mental effort, and to explore further questions around helpful abstractions to inform machine learning (some of which is described in the Impact section that follows). Hybrid Methodology is amenable to subsequent analyses that build on or depart from the initial analysis of the data, both because there are many "kinds" of data (e.g. quantitative, qualitative, self-reports, interpretation) to work with and because there are disciplinary experts who are already familiar with that data from the interdisciplinary work.

HYBRID METHODOLOGY IMPACT: GOING BACK TO OUR INTELLECTUAL COMMUNITIES WITH RELEVANT FINDINGS

Having multiple analytical entry points into a hybrid methodology data set can provide a team opportunity to make impact in a variety of ways and for different intellectual communities (both company-internal and external). The richness and variety of the hybrid methodology data set, and the analyses described above, left our team poised to develop work products (i.e. outputs, deliverables) that generated impactful early outcomes for context-aware assistive technology, including: (1) shaping early user experience design, (2) informing the research agenda for future studies in cognitive science, and (3) developing nascent research on infrastructure for assistive technologies. Together these follow-on projects represent a portfolio approach to delivering impact from a hybrid methodology data set, leveraging and extending the data and analysis in different ways.

Each of these three follow-on projects had distinct ambitions for how to deliver relevant findings to the "home discipline" intellectual communities that came together at the outset of our hybrid methodology project. The follow-on projects offer contrasting approaches to extending the analysis and application of a hybrid methodology data set, and suggest ways that qualitative data could be used in machine learning and cognitive science. The first two of our listed outcomes were in sense more straightforward or familiar. One involved envisioning a series of end-user design concepts based on the insights — means of interventions, broadly, that users might find helpful. The other involved addressing a single cognitive science research question emerging from the analysis of outlier mental effort scores (Jonker et al. in review).

This section focuses on the third on the list — a follow-on project focused on technology infrastructure development — to illustrate a form of impact that can be created through work products that may be novel in applied ethnography. This project involved developing and partially implementing two data labelling protocols based on abstractions deemed potentially useful for context-aware assistive technology. The abstractions, protocols, and resulting labelled data set each served an early informative role in infrastructure development.

Building frameworks or abstractions that make sense of the human, social world should feel familiar to applied ethnographers. Abstractions are also the foundation for making any machine learning possible. Without abstractions, machine learning models would have to cope with an infinite amount of categories with one data point each. For example, we might use the abstraction of a dog to build machine learning models that are able to detect dogs across breed, age, size, and so on. In our setting, the most useful level of abstraction would allow a machine learning model to reduce the inherent complexity of context and to hone in on what is most relevant for the human in a given moment.

To guide the development of useful abstractions, we studied the literature of conversational agents, or chatbots, an area where researchers have encountered similar challenges in terms of complexity. Our task involved us attempting to "read" and interpret a context for meaning. Similarly, chatbot-development involves seeking to extract "meaning" embedded in the syntax of language, treating a text as more than a sequence of words. Recent work has shown how hierarchies of abstractions can improve the performance of chatbots. In particular, research scientists Khatri et al. (2018) find that incorporating dialogue acts, inspired by philosopher John Searle (1969), can improve the performance of their contextual topic model for dialogue systems.

Inspired by recent advances in the field of conversational agents, we developed two sets of abstractions, Set A and Set B, that repackaged and represented the strongest patterns around the experience of context emerging from our hybrid analysis processes. Abstraction Set A was more holistic (more "zoomed out" in its representation of aspects of context) whereas Abstraction Set B was more granular, and broke down context into several components. Each abstraction set was mutually constitutive of the other (i.e. each abstraction set represented and reframed the content of the other), but each was also independent of the other (i.e. one set did not need the other set in order to be legible).

Our abstractions served as the foundation for the development of several data labelling protocols, which consisted of a set of instructions for how to generate a labelled data set of human experience of context. Ultimately, this labelled data set is needed to train a machine learning model to detect Abstraction Set A and B. In order for annotators to be able to label a piece of data as a given abstraction, they need to know what the abstraction is, which in our case was not as straightforward as, for instance, labelling whether or not an image contains a dog. Most of us share an understanding of the abstraction of a dog, and we have no difficulty pointing at examples. In comparison to the abstraction of a dog, Abstraction Set A and Set B were more ambiguous, and closer to concepts such as "freedom" or "democracy." There is a rich tradition in the social sciences for how to reliably encode data with abstract concepts. Political science, in particular, contains several examples such as the Polity data, which rates countries on a numeric scale from democratic to authoritarian, the Comparative Manifesto Project, containing coded summaries of political party manifestos, an-often used source for placing parties on a left-right scale, or Transparency International's yearly corruption perceptions index.

Traditionally, and in all three examples mentioned above, data that involves more abstract concepts are generated by experts, often academics with deep subject matter expertise. However, to generate data at sufficient scale to train a machine learning model, we need to be able to move beyond experts who are generally costly and in short supply. Thus, we needed to ensure annotators had a sufficiently nuanced understanding of our abstractions to be able to label data *as if* they were experts without requiring them to be trained ethnographers or have deep knowledge of our project — they abstractions needed to be teachable. Further, they needed to be able to detect an abstraction from video and audio alone, without access to our field notes. Despite the lower expertise of naive annotators, recent research indicates that deploying crowd-sourcing can generate results that are indistinguishable from expert approaches (see Benoit et al. 2016 for an example in the context of political texts).

Teaching new abstract concepts is hard. We took an examples-based approach, in which the abstractions were primarily taught through instructive examples in the form of brief video clips from our field recordings. We first provided the annotators a brief description of the abstraction. Afterwards, the annotators were shown three examples that highlight various aspects of the abstraction. The first example is a prototype of the given abstraction. This is the clearest illustration of the abstraction we have in our data. However, a clear example is not enough to be able to meaningfully label data from long-form video. It is equally important that annotators understand that moments can vary along important dimensions and still belong to the same abstraction. For this reason, we provide two additional examples that highlight meaningful variation within the abstraction. These examples helped annotators understand the different dimensions of an abstraction, which in turn helped them set boundaries and differentiate between various abstractions. Ideally we would have had a fuller training set, of several examples with a lot of variety for each abstraction, but at the time we had three training examples for each. Ultimately, the labelling protocol needs to strike a balance between sharing enough information to learn the abstraction but not so much information that the protocol starts to resemble an exhaustive catalog of variants.

Testing the annotators required a validation strategy. We were looking to test the degree to which the naive annotators were able to replicate our "expert" labels. To develop a benchmark, we took a piece of data and labelled it based on consensus among us as researchers over whether or not a given abstraction was present in the data. We did this for Abstraction Set A and Abstraction Set B, because we wanted to compare results and see which abstraction set was more easily learnable for naive annotators. Establishing this benchmark collaboratively, as an interdisciplinary team, meant an iterative discussion and refinement of what the definitions of the abstractions themselves were.

Training and testing annotators on the abstractions took a full day, and the data set available was modest (50 hours of video footage), as it came directly from the fieldwork. The team provided the training and assessed annotators as a group and individually against the benchmark data set, labelled by us. At this phase of the project the ambition was not to develop training data for a machine learning model, but to explore whether it was possible for a group of naive annotators to learn and apply our abstractions. Going forward, we envision a process where naive annotators are initially screened based on their ability to replicate our "expert" labels. After this screening, new data should be labelled based on majority voting among selected annotators, as is common in the literature (e.g., Fridman et al. 2019).

Our initial results are mainly positive. Overall, annotators had above-chance ability to agree with our labels, with the best-performing annotators missing the benchmark by only 5%. For the most part, there was relatively high intra-rater agreement between the naive annotators, indicating that different naive annotators would be able to independently reproduce approximately our abstraction labels. The team found that the best performing naive annotators understood and implemented the "rules" for coding (e.g. all abstractions in Abstraction Set A are mutually exclusive) and shared an understanding of the granularity of the labelling task. Difficulties in this area invited errors of two kinds: either parsing the long form video data too granularly (applying labels to less than salient evidence of an abstraction) or not granularly enough (failing to apply labels to salient evidence of an abstraction).

In our testing phase, we also captured, but have not yet analysed, annotators' certainty when labelling a given piece of data, as well as both point estimates and ranges of start and end times for a given abstraction. These data allow us to analyse accuracy across different levels of (un)certainty, and understand the degree to which annotators disagree about when an abstraction starts and ends. We also allowed annotators to tag and suggest new potential abstraction labels within Abstraction Set A or B (whichever one they were coding), as a way to generate new potential abstractions that could be refined in further analysis going forward.

This process of developing a data labelling protocol led to some overall lessons on making work products from hybrid ethnography that are relevant to specific intellectual communities. Applied ethnographers should have the ability to recognize the limits of work products and their utility — for instance, abstractions alone are not useful in building technology infrastructure. Indeed, applied ethnography often deals in abstractions or frameworks but does not often go a step further and apply them to machine learning problems — to do this ethnographers need to build another kind of work product, namely data labelling protocols. Developing data labelling protocols requires developing training material that links abstractions back to very concrete, detectable, recognizable examples in the data. This requires, at a certain point in a project, moving away from the nuance and complexity of ethnographic thinking and being quite firm and mutually exclusive about what something is or isn't in order for labelling to be possible. Establishing benchmarks (for the annotators to learn) requires consensus among the researchers, and iteration and refinement of the abstractions themselves in the process, as researchers are forced to be very clear about what something is or isn't. We have found that this process helps to sharpen the precision of the original concepts themselves. Labels must then be tested with annotators, who look at raw data and label based on the learned abstractions — and here what we discovered is that inter-rater reliability when dealing with such complex topics (human experience of context) might be lower than what is organizationally common or acceptable, and that annotator training for such complex topics is time-consuming and research-intensive.

Overall, if applied ethnographers want to influence infrastructures like those that support context-aware assistive technologies, these teams are greatly helped by a willingness to extend their frameworks and use them to form new work products. In our case of hybrid methodology we did so by extending our abstractions into a tested data labelling protocol, in addition to informing experience design and the research agenda for cognitive science teams.

HYBRID METHODOLOGY PROCESS: GUIDELINES FOR THE PROCESS OF INTERDISCIPLINARY TEAM COLLABORATION

Research that is both interdisciplinary and collaborative requires a balancing act between the practices of one discipline and another, such that the team develops new hybrid practices — this in turn means that working together is a process that cannot be taken for granted. The sections above have outlined the key hybrid methodology approaches for research, analysis, and impact. What follows are guidelines for effective collaboration within an interdisciplinary team, the order of the points organized by when in the project process that point is most relevant and useful (from framing to analysis), with the last point being about the general ethos throughout a project, based on our learnings.

Let the Research Question Be the Team's Home Base

For complex research questions, we need to flip the decision-making process on its head. Rather than using a discipline to define the methodology, we instead let the research question drive the methodology decisions. The major advantage of a highly interdisciplinary team is that it unlocks a large set of tools and methods that can be used to answer a central research question. We found that certain methods came to the fore at distinct stages of our research, and that each discipline had something crucial to contribute at different stages of the design and analysis, so we strove to set aside the mentality of "this is how we conduct research in Discipline X" and instead adopt the thinking, "this is how we best answer Question Y." The resulting process is more than interdisciplinary; the cross-pollination and switching between methods becomes so frequent and fluid as to create something more like a *hybrid* — hence hybrid methodology.

Prepare for an Immersion Into Each Other's Fields

Interdisciplinary projects work best when each discipline is given opportunity to contribute, but also when each discipline understands the other. This is not simply learning about each discipline's methodologies and problem-solving approaches, but deeply understanding their perspectives and world views. We would advocate for an early immersion, in which each disciplinary expert spends the day shadowing the other, trying to understand how each views the world. This entails listening and observing with openness —

what does the workflow for a machine learning engineer actually look like? How does a cognitive scientist run an experiment? How does an ethnographer conduct participant observation in the field? Each discipline expert should spend some time in the role of the other prior to fieldwork. When in the field, this spirit of immersion in each other's perspectives can continue by having researchers with different expertise gather data together. We agreed that two researchers, each from a different discipline, should go into the field together to meet with each participant. This setup gives researchers with a range of knowledge a shared perspective from which to draw — they can discuss how they, in pairs, observed or noted different aspects of the same context, having both been in the field.

Build in the Ability to Iterate Extensively

Interdisciplinary projects require constant developing and improving of approaches based on contributions across disciplines and shared learnings as a unit. We advocate for building in ways to iterate throughout the process. For example, data collection might be structured so that it occurs in two parts with a break in-between to assess and refine approaches and develop early insights. The team can then reconvene at the end of the second part of data collection to review the revised approaches and analyze the data. The discipline experts should regularly review and weigh in on analyses in progress. Time and logistics for this iteration should be built into the project timeline and scope — for instance, ensuring all experts have opportunities to meet and work together in real-time at key moments in the research when approaches are being built, assessed, or (if necessary) rebuilt. This may not be unique to hybrid methodology, but it is likely especially critical given the diversity of the research and researchers.

Work with Fuzzy Definitions and Cross-Disciplinary Translations

Language becomes especially important in interdisciplinary projects, as different disciplines might have different definitions of the same term (e.g., “context”) or terms might not yet exist for newly observed phenomena. It is vital to do translation exercises across disciplines, particularly with terms that are common among the disciplines but defined differently in each — for instance, how do machine learning concepts map onto anthropological concepts (e.g. “abstraction” and “pattern”), and how do cognitive science understandings of experience map onto phenomenological and philosophical understandings (e.g. emotion and effort)? In cases where a phenomenon is not well defined by either discipline, new language emerges. We found ourselves working with fuzzy definitions, making a point to talk about what we did not fully know yet, in an effort to define as we went along what these terms meant (for example, the terms we used to break down the components of context), and working toward more concreteness of terms over the course of the project.

Recognize the Value of Different Types of Data

“Data” is one of those terms that is common across disciplines and yet comes in unique forms, from pixels to 0s and 1s to the thick description of a wink (Geertz 1973). Interdisciplinary projects benefit from the full team re-defining “data,” such that each discipline feels that there is both familiar and unfamiliar data being captured. It is important to recognize the value in unfamiliar data and to recognize that data which feels unusable for one discipline is actually incredibly relevant in another**.** Many disciplines (anthropology, machine learning, cognitive science) value taking a data-driven approach, but that "data" itself may look very different for each discipline.

Find the Highest Helpful Level of Abstraction

In order for an insight or concept (e.g. about human experience or human behavior) to be relevant and actionable across disciplines, it needs to have a certain level of abstraction from raw data so that it translates not only across individual data points but across different disciplines, yet it cannot be so abstract that it loses too much specificity and actionability, rendering it meaningless for each discipline. In our case, abstractions ideally allow us to develop knowledge that generalizes beyond any one individual's experience of context, to allow for actionability or relevance beyond our participant pool. For example, it might be too abstract to say that social interactions are one aspect of context that impacts experience, but to say that certain types of social interactions (e.g. caretaking, collaboration) impact the experience of context might be at the "right" level of abstraction to be directive about what value to offer in interventions or how to build for those interventions. In our project we have learned the value of 'imperfectly useful abstractions' that helped us to generalize enough given we were addressing a technology that doesn't yet exist, and yet that required constant re-evaluation and adjustments to the granularity of the abstraction (similar to our points about fuzzy definitions and translations above). Abstractions help us to pinpoint relevancy. In the words of scientist and engineer Edsger W Dijkstra, "[...] the purpose of abstracting is not to be vague, but to create a new semantic level in which one can be absolutely precise" (Dijkstra 1972, 864).

Know When and How to Shift between Description and Interpretation

In our project, we constantly discussed toggling between "bottom-up" and "top-down" analysis — and essentially this was a discussion about when to dwell in description and when to dwell in interpretation. It has been vital for us to have a high degree of granularity in the data (knowing that the data itself takes various forms), and staying close to the data for perhaps longer than on other applied projects, before reaching conclusions. But it has also been vital for us to move towards interpretation perhaps sooner than felt comfortable in other traditional within-discipline approaches (because given the quantity and quality of data captured, and the unfamiliarity with some of the data, we could have stayed close to the data for a long time). Moving to interpretation of the data allows us to build initial ontologies and categories for how to sort and make sense of the data, tying it to clear implications for what it is we are trying to inform. What has been most vital is the *shifting* between description and interpretation and back again — once we have some potential interpretations, going back to the descriptions to re-evaluate and refine.

Know When and How to Shift Between *Talking About* Approaches for How to Do Work and *Using* Approaches To Do Work

A consequence of having a process that cannot be taken for granted is that the team must make deliberate decisions and reach consensus on what teams would otherwise intuitively dive straight into doing — and this takes time. For example, once a disciplinary team has its data, that team generally knows how to analyze that data; this was not the case with us. We spent a considerable amount of time discussing which analysis approaches we would need in order to answer our project's questions, debating the pros and cons of each approach (and in these discussions it can be initially difficult for value judgements to not come into play, particularly about what data or results should look like). While these discussions were certainly crucial, we had to learn when to stop talking and start doing (or trying-to-do), in order to achieve tangible results. In such interdisciplinary situations, deciding on an approach can seem scary and wrong — what if it turns out the approach doesn't work and ends up being a waste of time? But when it felt like the team had spent too much time on a "meta" discussion about what to do, we learned to time-box discussions and instead invest the time the team would have spent debating the approaches into instead testing one or two (even for just a couple of hours), then regrouping. The fruitfulness of an approach can sometimes only be assessed by giving it a try and looking at the results. Instead of resolving methods debates based on "best practice," interdisciplinary projects may need to resolve these debates based on "the shoe that fits."

Seek Out Methodological Bricolage

In all, we have learned that interdisciplinary projects require some discomfort and compromise. Methodologies and approaches require give-and-take — no methodology is going to work as neatly as it would in its home-discipline. The orientation of the group should be towards a methodological bricolage of sorts: melding together traditional approaches in untraditional ways to make something new. Each discipline should be constantly looking to the edges of the field (e.g. how can we ask for scores of people's mental effort in-the-moment that take into account the reasons why the scores were given? How can we break a moment phenomenologically down to a handful of seconds in collaboration with participants?). This approach ultimately pushes each discipline further, together.

DISCUSSION: CONTEXT-AWARE ASSISTIVE TECHNOLOGY, HYBRID METHODOLOGY, AND THE IMPACT OF ETHNOGRAPHERS

Context-Aware Assistive Technology

Hybrid methodology has proven useful in beginning to address the complex problem of understanding the individual experience of context for personal computing and assistive technology. For instance, the study's findings indicate that people's broader goals and their social context and relationships play a critical role in characterizing high mental effort, even more so than environmental and task-based context (Jonker et al. in review). From a practical standpoint, these findings identify the most worthwhile context factors to pursue in future cognitive science and machine learning research. Moreover, the study has helped create new terms (or abstractions) to define different experiences of context, and different components of context that become relevant to an individual. This has challenged the notion that context — in particular, mental effort in context — is only experienced in terms of highs and lows, more or less, good or bad. It has even challenged the assumption that mental effort is a singular construct — it may in fact be the case that there are several "flavors" of mental effort in the real world (Jonker et al. in review). A deeper understanding of context has sought to help inform some of the success criteria of context-aware assistive technology that does not yet exist yet — assistive technology that perhaps knows not only what to intervene with, when, and how, but also when *not* to intervene. There are many unanswered questions about how assistive technology can help, rather than hinder, how people want to act upon their world, but hopefully there is now also the beginning of a collaborative way to talk about those questions.

Hybrid Methodology

Hybrid methodology presents an opportunity (and challenge) for disciplines to move beyond comfort zones. For anthropologists, it can mean coming up with a theory for understanding very messy and complicated contexts in a way that yields insights relevant to machine learning and cognitive science. For cognitive scientists it can mean exploring how lab studies and field studies build on or supplement one another, and how isolated variables studied in a lab (such as cognitive load during a puzzle challenge) can be studied systematically in everyday contexts alongside a number of other variables (such as emotion or mind-wandering) to further inform an understanding of cognition. For data and computer scientists and engineers it can mean understanding how qualitative data might provide helpful abstractions that can uncover new value propositions for machine learning and feature engineering. Across disciplines, there is an opportunity and challenge to explore how qualitative and quantitative analyses can work together on a shared data set. We hope that future interdisciplinary teams (particularly teams that bring *new* disciplines into the mix beyond the ones here) develop new methods at the intersection of existing ones, and new ways of analyzing, and defining what constitutes as, data. We hope these teams develop new types of outcomes that are relevant and impactful in "home disciplines," and new processes for collaborating to best bring out what is both at the core and the cutting edge of each discipline.

Next Steps for Ethnographers

Ethnography, in theory, holds promise complementing the approaches of machine learning and cognitive science, and addressing the challenges inherent in highly-controlled lab settings because it is embedded in the everyday, complex, "messy" reality of human life. Ethnographers are experts of context, abstracting out from thick descriptions of individuals. An algorithmic model, too, needs to be able to generalize to similar contexts and similar groups of users. Ethnography could have the potential to provide useful abstractions, descriptions and re-descriptions of the data that can inspire machine learning scientists to engineer new features that they had not before considered. It could help engineers determine what data and sensors to prioritize from the end-user's perspective. Ethnography could also have the potential to both augment quantitative metrics on cognition (such as mental effort highs and lows) with qualitative descriptors, and help to record such measurements more seamlessly in naturalistic settings. This contribution is deeply valuable because knowing that metrics like mental effort are high or low does not do enough to inform the device of when and how to intervene, or if it should intervene at all. The device also needs to know *why and how* mental effort spikes or drops because of an individual's experience of context. Ethnography can perform the knowledge discovery to scope out a space for future data collection and machine learning.

But ethnography, in practice, has yet to truly integrate into the early development of how these ubiquitous technologies work — both their ability to parse context and their ability to support human cognition. User research and qualitative data are typically part of defining "what we build" while machine learning and cognitive science are typically part of defining "how we build" — and there is little collaboration. This setup works well enough when the machine learning researchers know which data they will need to use for more constrained problems and use cases, but in the enormous complexity of everyday contexts (i.e. "the real world"), ethnographers can generate data, insights, and deliverables that help to define and scope machine learning work and bring qualitative insights early into the shaping of technologies and capabilities that do not exist yet. This requires that ethnographers roll up their sleeves, understand new emerging spaces, dive deeply and openly into new disciplines, and adaptively build a hybrid methodology around emerging research questions. It requires rethinking ethnographic research and outputs, and making these understandable and relevant to collaborator-disciplines. Although it is a challenge, the applied ethnographers who are willing to take it on may find themselves contributing to the definition of the next wave of ubiquitous computing, and in the process pushing the boundaries of ethnography's methods and applications.

Maria Cury is a manager at ReD Associates. Currently Maria studies technology in daily life to advise on product development, and is interested in advancing applied ethnographic research methods. Maria received an MSc in Visual, Material, and Museum Anthropology from Oxford, and a BA in Anthropology with Visual Arts certificate from Princeton University. mcu@redassociates.com

Eryn Whitworth is a post-doctoral research scientist at Facebook Reality Labs. Currently, Eryn is focused on advancing the practice and discourse in product user experience research through the development of new data representations and data sets depicting mundane activity. She received a PhD in information studies from the University of Texas at Austin. eryn@fb.com

Sebastian Barfort is a data scientist at ReD Associates. Sebastian works with clients in technology, financial services and healthcare to translate ethnographic insights into algorithms. Sebastian received a PhD in behavioral economics from the University of Copenhagen and double master's degrees from London School of Economics and NYU. *sebastianbarfort@gmail.com*

Séréna Bochereau is a technical program manager at Facebook Reality Labs. Séréna has a PhD in Haptics from Sorbonnes Universities and an MEng in Materials Science from University of Oxford. sbochereau@fb.com

Jonathan Browder is a research scientist at Facebook Reality Labs. His research interests include multisensory perception, human behavior in augmented and virtual reality, and non-parametric modeling. He received a Ph.D. and MA from Washington University in St Louis and a BA from Washington and Lee University, all in mathematics. jonathan.browder@oculus.com

Tanya Jonker is a research scientist at Facebook Reality Labs. Her work focuses on interaction between futuristic technologies and human cognition. She is currently exploring input and interactions with augmented and mixed reality, and how these devices might enable new types of cognitive offloading. Tanya received a PhD in Cognitive Psychology from the University of Waterloo. Tanya.Jonker@oculus.com

Sophie Kim is a UX research scientist in Facebook Reality Labs at Facebook. Sophie focuses on bringing human-centered and experience-driven approaches to future-facing research and development. She has a special interest in augmented reality interactions and how ethnographic research can help inform it. Sophie received a PhD in Human Factors Engineering from Virginia Tech. sophiekkim@fb.com

Mikkel Krenchel is the Director of ReD Associates North America. He has spent a decade advising leaders across a wide range of Fortune 500 companies on corporate and product strategy, and led ReD's emerging practice for integrating social and data science. mkr@redassociates.com

Morgan Ramsey-Elliot is a partner at ReD Associates, where he works with technology, financial services, and retail companies. He enjoys working at the intersection of "old" and "new," advising on product strategy for both well-established companies striving to adapt to the digital economy, and digital-native companies growing into maturity. mre@redassociates.com

Friederike Schüür, PhD is a data and machine learning scientist. She leads machine learning efforts at Cityblock Health, serves on the data advisory board of USA for UNHCR, and she is a long-standing data science for social good volunteer with DataKind. She loves data in all its shapes and sizes. friederike.schueuer@gmail.com

David Zax is a senior consultant at ReD Associates, where he has focused on conducting ethnographic research for tech companies. Prior to ReD he was a freelance journalist contributing to Fast Company, Technology Review, This American Life, and The New York Times. Contact: David.zax@gmail.com

Joanna Zhang is a senior researcher at ReD Associates. In past lives, she's waitressed in NYC, coached high school debate, organized public health campaigns, designed for an architecture studio, supported digital strategy at the White House, and dipped potato chips in chocolate by hand at a candy/hotdog shop. jze@redassociates.com

NOTES

Acknowledgments – The authors would like to thank Sean Keller, Joshua Walton, Jerome Holman, James Hillis, Sean Trowbridge, Karl Ridgeway, and Hrvoje Benko for the support of this work at Facebook Reality Labs as well as Dave Baiocchi, David Kearford, Michael Marcusa, Mads Nielsen, and Gumbo for their support with this work at ReD Associates.

REFERENCES CITED

anderson, ken, Dawn Nafus, Tye Rattenbury. 2009. "Numbers Have Qualities Too: Experiences with Ethno-Mining." *Ethnographic Praxis in Industry Conference Proceedings:* 123-140.

Arora, Millie P., Mikkel Krenchel, Jacob McAuliffe, and Poornima Ramaswamy. 2019. "Contextual Analytics: Towards a Practical Integration of Human and Data Science Approaches in the Development of Algorithms." *Ethnographic Praxis in Industry Conference Proceedings:* 224-244.

Benoit, Kenneth, Drew Conway, Benjamin E. Lauderdale, Michael Laver, and Slava Mikhaylov. 2016. "Crowd-sourced text analysis: Reproducible and agile production of political data." *American Political Science Review*, *110* (2): 278-295.

Biehl, João. 2013. "Ethnography in the Way of Theory." *Cultural Anthropology* 28 (4): 573-597.

Bornakke, Tobias and Brian L. Due. 2018. "Big–Thick Blending: A Method for Mixing Analytical Insights from Big and Thick Data Sources." *Big Data & Society* January—June 2018: 1-16.

Braverman, Irwin. 2011. "To see or not to see: How visual training can improve observational skills." *Clinics in Dermatology* 29 (3): 343-346.

Christensen J. C., Estepp J. R., Wilson G. F., and Russell C. A. 2012. "The Effects of Day-to-Day Variability of Physiological Data on Operator Functional State Classification." *Neuroimage* 59: 57–63.

Hwan-Hee Choi, Jeroen J. G. van Merriënboer and Fred Paas. 2014. "Effects of the Physical Environment on Cognitive Load and Learning: Towards a New Model of Cognitive Load." *Educational Psychology Review* 26 (2): 225-244.

Csikszentmihalyi, Mihaly. 2008. *Flow: The Psychology of Optimal Experience.* New York: Harper Perennial Modern Classics.

Cury, Maria and Daniel Bird. 2016. "Applying Theory to Applied Ethnography." *Ethnographic Praxis in Industry Conference Proceedings:* 201-216.

Dijkstra, Edsger W. "The Humble Programmer." 1972. *ACM Turing Award Lecture* 15 (10): 859-866.

Duranti, Alessandro and Charles Goodwin. 1992. "Rethinking Context: An Introduction." In *Rethinking Context: Language as an Interactive Phenomenon,* 1-42, Vol. 11, edited by Alessandro Duranti and Charles Goodwin. Cambridge: University Press.

Engeström, Yrjö, Reijo Miettinen, and Raija-Leena Punamaki (eds.) 1999. *Perspectives on Activity Theory*, Cambridge: University Press.

Fraser, Kristin, Irene Ma, Elise Teteris, Heather Baxter, Bruce Wright, and Kevin McLaughlin. 2012. "Emotion, Cognitive Load and Learning Outcomes During Simulation Training." *Medical Education* 46: 1055-1062.

Fridman, Lex, Daniel E. Brown, Julia Kindelsberger, Linda Angell, Bruce Mehler, and Bryan Reimer. "Human Side of Tesla Autopilot: Exploration of Functional Vigilance in Real-World Human-Machine Collaboration." 2019.

Geertz, Clifford. 1973. *The Interpretation of Cultures.* New York: Basic Books.

Geertz, Clifford. 1998. "Deep Hanging Out." *New York Review of Books* 22 October 1998 issue. Accessed November 1, 2019. https://www.nybooks.com/articles/1998/10/22/deep-hanging-out/

Glaser, Barney G. and Anselm L Strauss. 2017. *Discovery of Grounded Theory: Strategies for Qualitative Research.* New York: Routledge.

Goldhagen, Sarah Williams. 2017. *Welcome to Your World: How the Built Environment Shapes Our Lives.* New York: HarperCollins.

Grasseni, Cristina, ed. 2007. *Skilled Visions: Between Apprenticeship & Learning.* New York and Oxford: Berghahn Books.

Helton, William S. and Katharina Näswall. 2015. "Short Stress State Questionnaire: Factor Structure and State Change Assessment." *European Journal of Psychological Assessment* 31: 20-30.

Hou, Carolyn and Mads Holme. 2015. "From Inspiring Change to Directing Change: How Ethnographic Praxis Can Move beyond Research." *Ethnographic Praxis in Industry Conference Proceedings:* 190-203.

Ingold, Tim. 1993. "The Temporality of the Landscape." *World Archaeology* 25 (5): 152-174.

Jonker, Tanya R., Eryn Whitworth, Kahyun Sophie Kim, Jonathan Browder, Serena Bochereau, Hrvoje Benko, Sean Keller, Sebastian Barfort, Maria Cury, Mikkel Krenchel, Morgan Ramsey-Elliot, Friederike Schuur, David Zax, Joanna Zhang. In review. "Mental Effort During At-Home Activities: Moments of High Mental Effort Can Reveal New Opportunities for Technology." Submitted to *ACM CHI Conference on Human Factors in Computing Systems 2020.*

Kane, Michael J., Leslie H. Brown, Jennifer C. McVay, Paul J. Silvia, Inez Myin-Germeys, and Thomas R. Kwapil. 2007. "For Whom the Mind Wanders, and When: An Experience-Sampling Study of Working Memory and Executive Control in Daily Life. *Psychological Science* 18 (7): 614-621.

Kaptelinin, Victor and Bonnie A. Nardi. 2006. *Acting with Technology: Activity Theory and Interaction Design.* Cambridge: The MIT Press.

Khatri, Chandra, Rahul Goel, Behnam Hedayatnia, Angeliki Metanillou, Anushree Venkatesh, Raefer Gabriel, and Arindam Mandal. 2018. "Contextual Topic Modeling For Dialog Systems." In *2018 IEEE Spoken Language Technology Workshop (SLT),* 892-899.

Killingsworth, Matthew A., and Daniel T. Gilbert. 2010. "A Wandering Mind is an Unhappy Mind." *Science* 330.6006: 932.

Larson, Reed, and Mihaly Csikszentmihalyi. 2014. "The Experience Sampling Method." In *Flow and the Foundations of Positive Psychology,* 21-34. Dordrecht: Springer.

Lottridge, Danielle, Mark Chignell, and Aleksandra Jovicic. 2011. "Affective Interaction: Understanding, Evaluating, and Designing for Human Emotion. *Reviews of Human Factors and Ergonomics* 7 (1): 197-217.

Madsbjerg, Christian and Mikkel Rasmussen. 2014. *The Moment of Clarity: Using the Human Sciences to Solve Your Biggest Business Problems.* Boston: Harvard Business Review Press.

Matthews, Gerald, James Szalma, April Rose Panganiban, Catherine Neubauer, and Joel S. Warm. 2013. "Profiling Task Stress with the Dundee Stress State Questionnaire." In *Psychology of Stress*, 49-91, edited by Leandro Cavalcanti and Sofia Azevedo. Nova Science Publishers, Inc.

Mauss, Marcel. 2009. *Manual of Ethnography.* Translated by N.J. Allen. New York: Berghahn Books.

Menkedick, Sarah. 2018. "Behind the Writing: On Interviewing." *Longreads,* July. Accessed November 1, 2019. https://longreads.com/2018/07/20/behind-the-writing-on-interviewing/

Nardi, Bonnie A., ed. 1996. *Context and Consciousness: Activity Theory and Human Computer Interaction*. Cambridge: The MIT Press.

Paas, Fred G. 1992. "Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive-Load Approach." *Journal of Educational Psychology* 84 (4): 429-434.

Pink, Sarah. 2007. *Doing Visual Ethnography.* London: Sage Publications.

Pink, Sarah. 2009. *Doing Sensory Ethnography.* London: Sage Publications.

Redelmeier, Donald A. and Daniel Kahneman. 1996. "Patients' Memories of Painful Medical Treatments: Real-Time and Retrospective Evaluations of Two Minimally Invasive Procedures." *Pain* 66 (1): 3-8.

Reis, Harry T. and Gable, Shelly L. 2000. "Event-sampling and Other Methods for Studying Everyday Experience. In *Handbook of Research Methods in Social and Personality Psychology,* 190-222, edited by H. T. Reis & C. M. Judd. New York: Cambridge University Press.

Roth, Wolff-Michael. 2004. Activity Theory and Education: An Introduction. *Mind, Culture, and Activity,* 11 (1): 1-8.

Schmeck, Annett, Maria Opfermann, Tamara van Gog, Fred Paas, and Detlev Leutner. 2015. "Measuring Cognitive Load with Subjective Rating Scales During Problem Solving: Differences Between Immediate and Delayed Ratings." *Instructional Science* 43: 93-114.

Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language.* Cambridge University Press.

Smallwood, Jonathan and Jonathan W. Schooler. 2015. "The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness." *Annual Review of Psychology* 66: 487-518.

Sweller, John. 1993. “Some Cognitive Processes and Their Consequences for the Organization and Presentation of Information.”*Australian Journal of Psychology* 45: 1–8.

Van Gog, Tamara, Femke Kirschner, Liesbeth Kester, and Fred Paas. 2012. “Timing and Frequency of Mental Effort Measurement: Evidence in Favor of Repeated Measures.” *Applied Cognitive Psychology* 26: 833–839.